COMPARISON OF LEAD CONCENTRATIONS IN THE TOP 1 INCH OF SOIL TO CONCENTRATIONS MEASURED IN SURFACE SCRAPE SAMPLES

HERCULANEUM LEAD SMELTER
HERCULANEUM, MISSOURI

CERCLIS ID: MOD006266373

Superfund Technical Assessment and Response Team (START) 3 Contract
Contract No. EP-S7-06-01, Task Order 0021

Prepared For:

U.S. Environmental Protection Agency
Region 7
Superfund Division
901 North 5th Street
Kansas City, Kansas 66101

September 20, 2007

Prepared By:

Tetra Tech EM Inc.
415 Oak Street
Kansas City, Missouri 64106
816-412-1741

INTRODUCTION

Tetra Tech EM Inc. (Tetra Tech) was tasked by the U.S. Environmental Protection Agency (EPA) Region 7 Enforcement/Fund Lead Removal program to compare, statistically, lead concentrations collected from the top 1 inch of soil with lead concentrations in surface scrape samples at selected locations within Herculaneum, Missouri (City). Specifically, EPA requested the Tetra Tech Superfund Technical Assessment and Response Team (START) to determine whether concentrations of lead were statistically significantly higher (or lower) in data sets comprised of samples collected within the top 1 inch of soil versus scrape samples from the soil surface. The assessment was conducted under the authority of the Comprehensive Environmental Response, Compensation, and Liability Act of 1980 and the Superfund Amendments and Reauthorization Act of 1986. The project was assigned under START Contract No. 68-S7-01-41, Task Order No. 0021.

Tetra Tech conducted this analysis using the data set called "Sanitized 7003 AOC Results." The data used for this analysis consisted of matched pairs of analytical results from each depth interval for the following groups of locations: quadrant samples from individual properties (Q1, Q2, Q3, and Q4), haul route (HR) samples, and samples from play areas (PA). No matched results were available for samples collected from gravel drives (GD) or gardens (GAR). The analysis was conducted for both the full data set (293 matched pairs) and a modified data set (266 matched pairs). The modified data set excluded the following samples with higher levels of historical contamination that were judged inconsistent with recent surface recontamination trends: EPA ID Numbers 54, 56, 148, 301 (quadrants 3 and 4 only), 485, 551, and 566. The two data sets will hereafter be referred to as the "full" and "reduced" data sets.

The statistical approach and results from this analysis are presented below. Section 1.0 provides a more formal and detailed description of the statistical methods used, while Section 2.0 provides a general description considered more suitable for presentation to audiences without a formal background in statistics.

STATISTICAL METHODS, RESULTS, AND CONCLUSIONS (Formal Presentation)

Lead results for samples collected from the top 1 inch and surface scrapings of soil were analyzed using Version 5 of the JMP® statistical software package from SAS® Institute. Because samples were collected from two depths at the same set of physical locations, the results were amenable to analysis using a matched-pairs statistical design. Statistical comparisons were conducted using both parametric (paired-difference *t* test, which assumes the paired differences between the two populations being compared follow a normal distribution) and nonparametric (Wilcoxon signed-rank test, which assumes

that the distribution of paired differences is symmetrical about the median) tests. Results were evaluated at the 5 percent ($p \leq 0.05$) level of significance (i.e., equivalent to a 95 percent confidence level). Additional details of the statistical tests are provided in EPA (2006), as well as in mainstream statistical texts (Zar 1999).

Results of the statistical comparison for the full and reduced data sets are presented in Figures 1 and 2, respectively. Figures 1 and 2 provide several graphical presentations of the data, as well as results of the formal statistical tests. Interpretation of the statistical output in Figures 1 and 2 depends on the question addressed in the test or, stated more formally, on the specific form of the null (H_0) and alternative hypotheses (H_A). For comparing measures of central tendency, one can just ask whether the mean (or median) difference in the pair-wise concentrations in the two depth intervals is equal to zero. This is stated as a two-sided hypothesis, and the form of H_0 and H_A are shown below:

Two-sided test – H_0: the mean (median) difference in the pair-wise results for the 1-inch and scrape samples is equal to zero

H_A: the mean (median) difference in the pair-wise results for the 1-inch and scrape samples is not equal to zero

For the two-sided test, there is interest only in whether the mean (median) difference is zero. This is mathematically equivalent to testing whether the mean concentrations are different, but is more appropriately defined as the mean difference within the context of the paired-difference test. If H_0 is rejected, then it is concluded that the mean (median) difference is not equal to zero, but there is no interest in further investigating which of the two groups of samples has the higher (or lower) mean (or median) concentration.

The parametric test results for both the full data set (n= 293, p= 0.06, Figure 1) and reduced data set (n= 266, p= 0.23, Figure 2) indicate that the mean differences in lead concentration are not statistically different from zero (i.e., mean concentrations in the 1-inch and surface scrape samples are not statistically different). The nonparametric test result also indicates that the median difference in concentrations is not significantly different from zero for the full data set (n= 293, p= 0.09, Figure 1). However, the nonparametric test result for the reduced data set

(n= 266, p= 0.001, Figure 2) indicates that the median difference in concentrations is significantly different from zero.

In order to resolve this contradictory finding for the two-sided parametric and nonparametric tests for the reduced data set, it is necessary to: (1) evaluate the assumptions of the paired-difference *t* test to

determine if parametric testing is appropriate, and (2) recognize the different nature of the two tests (i.e., test of mean versus median difference in concentrations). A primary assumption of the parametric test is that the pair-wise differences in concentrations across all locations are normally distributed. This was tested using the Shapiro-Wilk W test at the 5-percent level of significance. The results of the Shapiro-Wilk W test are provided in Figures 1 and 2, and show that at the 95 percent confidence level, the paired differences in concentration are not normally distributed. The parametric test is considered to be robust to moderate departures from the assumption of normality, but is not robust to the presence of outliers (i.e., extreme differences in concentration). Therefore, results of the nonparametric Wilcoxon signed-rank test were judged to be more appropriate for comparing differences in concentration.

It is important to understand the subtle differences between the parametric and nonparametric tests. The parametric test evaluates the mean difference in concentrations, and is affected by extreme differences in concentration between the 1-inch and scrape results at each sampling location (e.g., differences were as large as 9,757 milligrams per kilogram [mg/kg] and 2,151 mg/kg, respectively, in the full and reduced data sets). The nonparametric test calculates the absolute value of the paired differences at each location, and then ranks the differences from smallest to largest. The median (mid-point for the ranked differences) difference in concentrations is evaluated in the nonparametric test, and this measure is relatively insensitive to the extreme differences in concentration that can confound interpretation of the parametric tests.

If there is interest beforehand in knowing the direction of any potential difference in concentration between the 1-inch and scrape samples, then a one-sided test is the more appropriate form. The hypotheses tested for the one-sided test are stated as follows:

One-sided test – H_0: the mean (median) difference in concentration between the 1-inch and scrape samples is less than or equal to zero

H_A: the mean (median) difference in concentration between the 1-inch and scrape samples is greater than zero

For the same significance (or confidence) level, the one-sided hypothesis test is said to have greater power compared to the two-sided test, which is a measure of the probability or likelihood that H_0 will be rejected when false. Because the assumptions of the parametric tests were not met (see discussion for the two-tailed test), only the results for the nonparametric tests are discussed below.

For the nonparametric comparisons, use of the one-sided hypothesis test lead to rejection of H_0 for both the full ($n= 293$, $p= 0.045$, Figure 1) and reduced ($n= 266$, $p<0.001$, Figure 2) data sets, leading to the

conclusion that the median concentration (i.e., based on measurement of the median difference in all pair-wise concentrations) of lead is higher in the scrape samples for both data sets.

APPROACH, RESULTS, AND CONCLUSIONS (General Presentation)

Lead results for samples in the full (n= 293) and reduced (n= 266) data sets collected from the top 1 inch of soil and from surface scrapings of soil were compared using graphical and statistical methods.

Figure 3 presents plots of the differences in concentration (1-inch result minus result for scrape sample) for all pairs of measurements in the full (top panel) and reduced (bottom panel) data sets. Positive differences shown in the plots in Figure 3 indicate that the 1-inch results are higher than the results for the scrape sample, while negative differences indicate that the results for the scrape samples are higher.

For the full data set, a total of 128 out of 293 or 44 percent of the results were higher in the 1-inch interval, with a maximum difference of 9,757 mg/kg. A total of 163 out of 293 or 56 percent of the results (two results were the same in both intervals) were higher in the surface scrape samples, with a maximum difference of 1,557 mg/kg. The average difference measured across all 293 samples was 90 mg/kg (on average, results from the 1-inch interval were 90 mg/kg higher than results for the scrape samples), but the median difference was -7 mg/kg (indicating that the median concentration was higher in the scrape samples).

For the reduced data set, a total of 108 out of 266 or 41 percent of the results were higher in the 1-inch interval, with a maximum difference of 2,151 mg/kg. A total of 155 out of 266 or 58 percent of the results (three results were the same in both intervals) were higher in the surface scrape samples, with a maximum difference of 1,557 mg/kg. The average difference measured across all 266 pairs of samples was -24 mg/kg (on average, results from the scrape samples were 24 mg/kg higher than results for the 1-inch samples), and the median difference was -14 mg/kg.

Results for the 1-inch and scrape samples from both data sets were compared using statistical tests appropriate when data represent a series of matched pairs (i.e., a result is available for both the 1-inch interval and from a surface scraping for all locations). The goal of the tests was to determine whether, on average, the reported results in the 1-inch samples were higher (or lower) than the results for samples collected from surface scrapings. Both parametric and nonparametric statistical tests were conducted to determine if the average difference across all pairs of results was significantly different from zero (an average difference of zero would indicate no difference between the 1-inch and scrape results). It was determined that a key assumption for the parametric tests (that the distribution of all pair-wise differences

in concentration follow a normal or Gaussian distribution) was not met in either the full or reduced data set and, therefore, that the nonparametric test results were more appropriate. The nonparametric test examines whether the median difference in concentrations (mid-point of the ranked or ordered concentration differences) is equal to zero, and does not assume that the differences follow a normal distribution.

The test results indicated with 95 percent confidence that, on average, lead concentrations were higher in the scrape samples for both the full (n= 293) and reduced (n= 266) data sets. This conclusion was based on testing whether the median (rather than the mean) difference in concentration across all pair-wise results was statistically different from zero.

REFERENCES

U.S. Environmental Protection Agency (EPA). 2006. "Data Quality Assessment: Statistical Methods for Practitioners, EPA QA/G-9R," EPA/240/B-06/003. Office of Environmental Information, Washington, DC. February.

Zar, J. H. 1999. *Biostatistical Analysis*, 4th Edition. Prentice-Hall, Inc. Upper Saddle River, NJ. 931 pp.

Figure 1 **Results of Matched-Pairs Statistical Analysis Comparing Lead Concentrations Measured in 293 1-Inch Soil and Surface Scrape Samples**

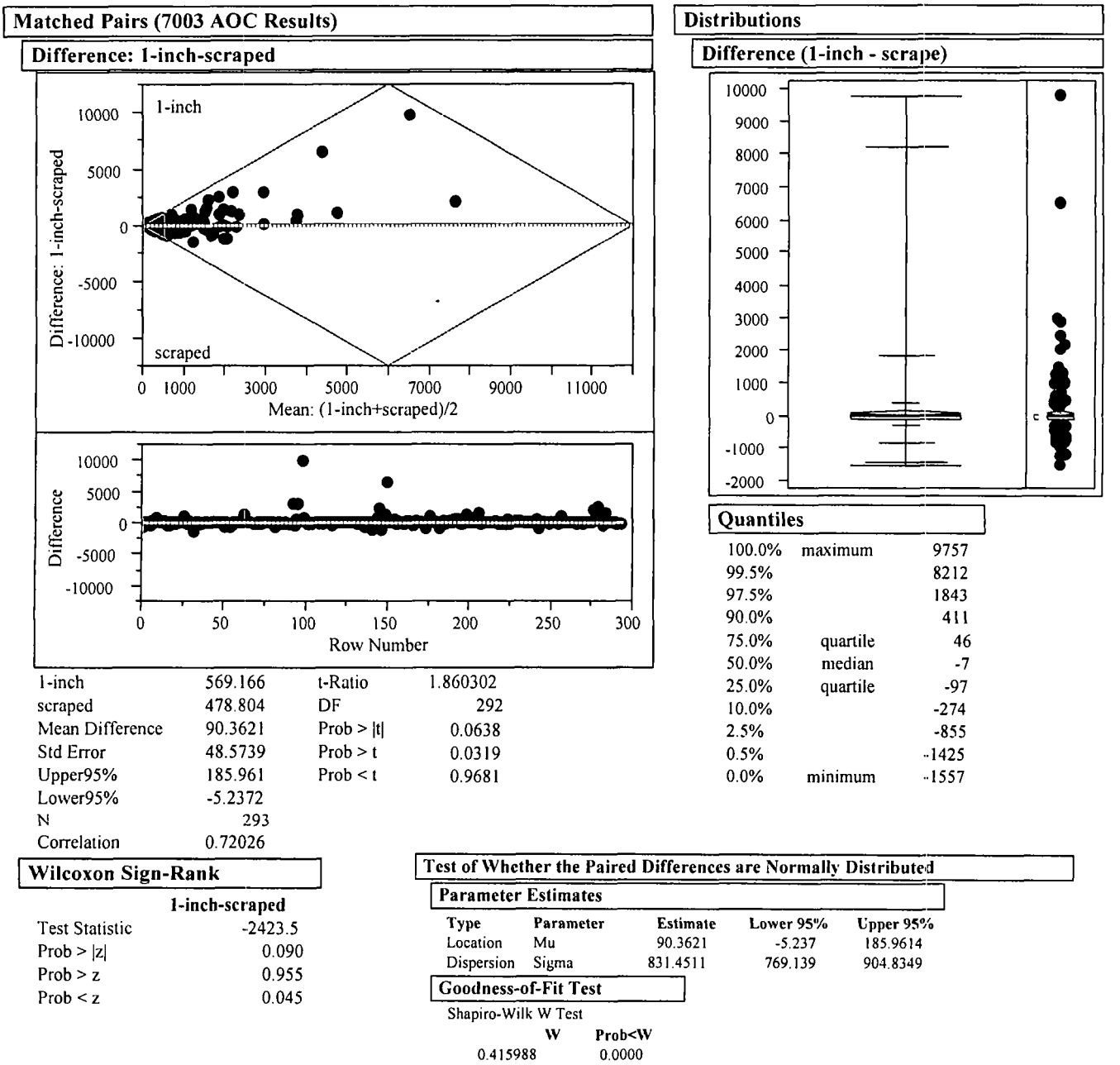


Figure 2 Results of Matched-Pairs Statistical Analysis Comparing Lead Concentrations Measured in 266 1-Inch Soil and Surface Scrape Samples

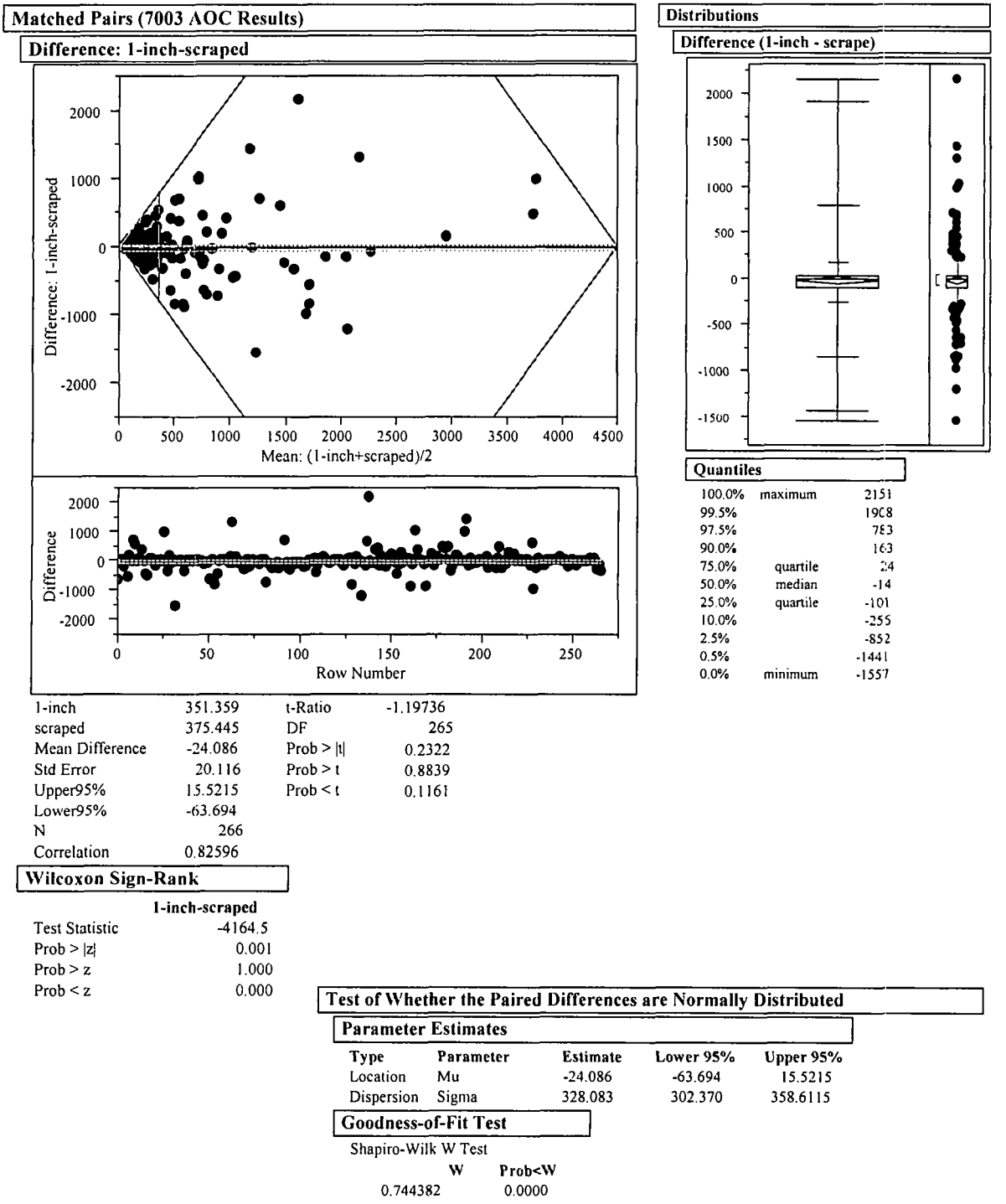
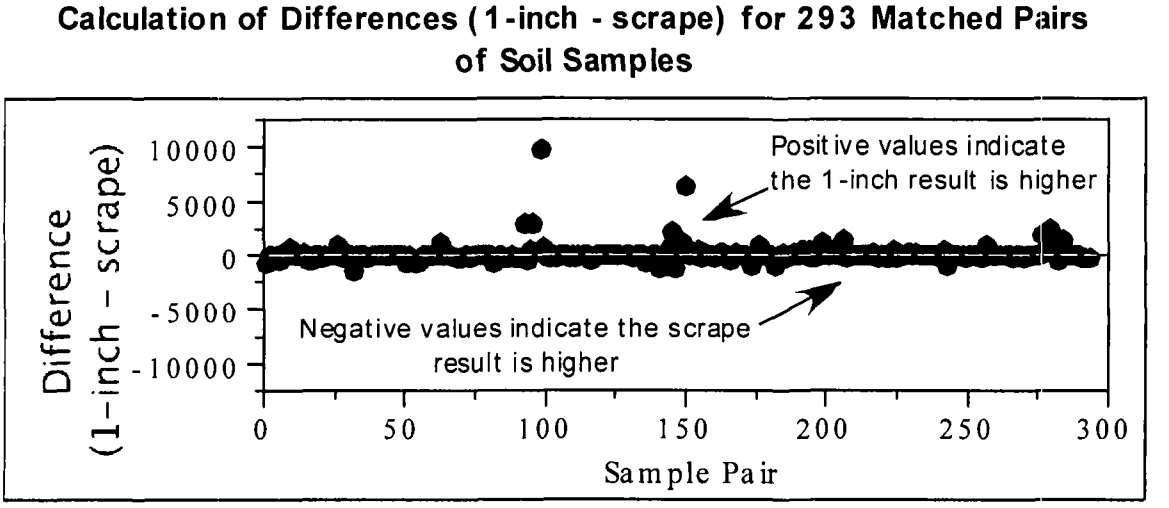
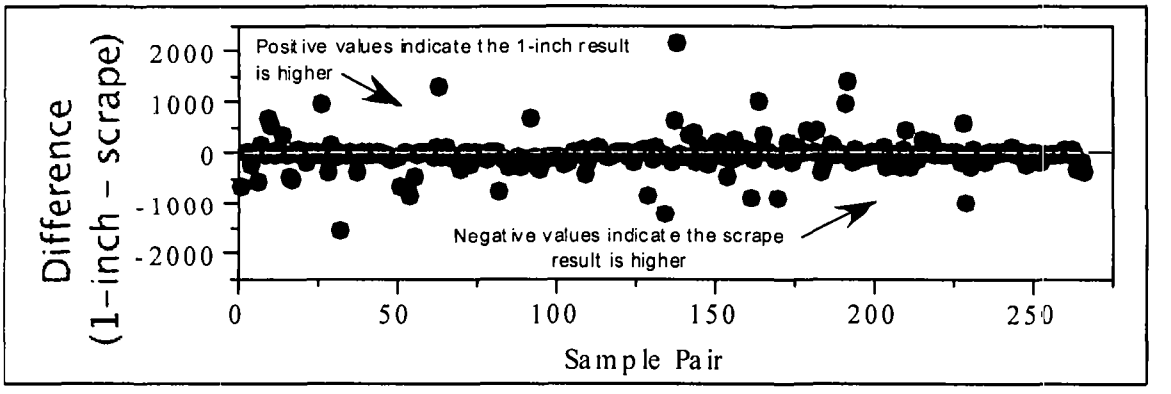


Figure 3 **Summary of Differences in Lead Concentrations Measured in Samples Collected from the Top 1 Inch of Soil and From Surface Scrapings of Soil for the Full (n= 293) and Reduced (n= 266) Data Sets**





TETRA TECH

September 20, 2007

Mr. Roy Crossland
START Project Officer
U.S. Environmental Protection Agency, Region 7
901 North 5th Street
Kansas City, Kansas 66101

Subject: **Comparison of Lead Concentrations in the Top 1 Inch of Soil to Concentrations Measured in Surface Scrape Samples**
Herculaneum Lead Smelter, Herculaneum, Missouri
U.S. EPA Region 7 START 3, Contract No. EP-06-01, Task Order No. 0021
Task Monitor: Bruce Morrison, On-Scene Coordinator

Dear Mr. Crossland:

Tetra Tech EM Inc. is submitting the attached Comparison of Lead Concentrations in the Top 1 Inch of Soil to Concentrations Measured in Surface Scrape Samples for the Herculaneum Lead Smelter. Using data provided by U.S. Environmental Protection Agency (EPA) from the site, Tetra Tech conducted a statistical analysis of the data to determine if lead concentrations in paired samples of the top 1 inch of soil differ significantly from those of scrape samples. Lead concentrations in the scrape samples were found significantly higher than those in the 1-inch samples.

If you have any questions or comments, please contact the project manager at (816) 412-1762.

Sincerely,

David Homer, PhD
Project Manager

Ted Faile, PG
START Program Manager

cc: Bruce Morrison, EPA
Ray Bienert, Tetra Tech

Enclosures

Tetra Tech EM Inc.
415 Oak Street. Kansas City, MO 64106
Tel 816.412.1741 Fax 816.410.1748 www.tetratech.com

X9004.06.0021.000